

## DOCUMENT RESUME

ED 194 567

TM 800 696

AUTHOR Beebe, Robert J.  
TITLE Use of the Behaviorally Anchored Rating Scale in Evaluating Teacher Performance.  
PUB DATE 80  
NOTE 21p.; Paper presented at the Annual Conference of the Evaluation Network (6th, Memphis, TN, September 30-October 2, 1980).  
EDRS PRICE MF01/PC01 Plus Postage.  
DESCRIPTORS \*Behavior Rating Scales; \*Evaluation Criteria; \*Personnel Evaluation; Special Education; Teacher Behavior; \*Teacher Evaluation; \*Test Construction; Testing Problems  
IDENTIFIERS \*Behaviorally Anchored Rating Scales

## ABSTRACT

Behaviorally anchored rating scales (BARS), a new quantitative method of employee performance evaluation, is advocated for teacher evaluation. Development of a BARS consists generally of five steps: a representative sample of potential raters generates the scales; the group identifies the broad qualities to be evaluated; the group formulates general statements representing various levels of performance for each quality and lists examples of behaviors at each level; a second group, representing the ratees, receives lists of the broad dimensions and of the specific behaviors and assigns the behaviors to the dimensions; raters then assign unused behaviors to describe satisfactory and unsatisfactory employees across several dimensions. A discrimination index is computed for each behavior and point values assigned to each behavior within the dimensions. Items above a criterion level of point variability are eliminated. Problems may arise with the raters' predictive acumen or possibly with the permanent assignment of values for specific behaviors. As an example of a BARS used in education, the scale developed by Marianne Price for evaluating special education teachers in the Montgomery County, Pennsylvania, Intermediate Unit, is appended. This scale measures professional growth, communications skills, rapport with students, planning, instructional skills, recordkeeping, behavior management, and interstaff relationships. (MH)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED194567

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
NATIONAL INSTITUTE OF  
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

USE OF THE BEHAVIORALLY  
ANCHORED RATING SCALE  
IN EVALUATING TEACHER  
PERFORMANCE

Robert J. Beebe

The University of Mississippi

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

R. Beebe

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

Presented at the Sixth Annual Evaluation Network  
Conference - Memphis, TN, September 30-October 2, 1980

TM 800696

## ABSTRACT

The paper addressed the concept, features, format, development, and relative merits of using the behaviorally anchored rating scale in evaluating teacher performance.

The discussion provides basic information on a technique which is at the cutting edge of industrial performance evaluation. Application of the behaviorally anchored rating scale technique to teacher performance evaluation is illustrated.

Suggestions as to how such a technique might be applied in a school district are offered, as well as critical comments intended to keep the technique in perspective.

## USE OF THE BEHAVIORALLY ANCHORED RATING SCALE IN EVALUATING TEACHER PERFORMANCE

As you are aware, numerous techniques have been developed, and are in current use, for the evaluation of employee job performance. There are various ranking procedures, including straight ranking, alternative ranking, paired comparison, and forced distribution techniques. There are qualitative methods, including critical incident, weighted checklist, and forced choice techniques. There are management by objectives approaches. There are also direct indexes, including measures of productivity, and measures of withdrawal, such as absenteeism and turnover.

By far the most popular methods of employee performance evaluation are those quantitative methods known as rating methods. [For a fine discussion and comparison of the various types of performance evaluation techniques, see Cummings and Schwab, 1973.] In recent years a new quantitative method of employee performance evaluation has developed, and has attracted considerable attention in the literature. This is the behaviorally anchored rating scale, or BARS for short. In the years since 1963, when the concept was first introduced, more than 125 journal articles have appeared in which the authors discuss, develop, test, or advocate the use of behaviorally anchored rating scales in evaluating employee job performance. (See Schwab and Heneman 1975; also DeCotiis, 1978).

Given the pressures both from within and without school districts for reliable and valid teacher performance appraisal mechanisms, it may surprise you to learn that only one application of the behaviorally anchored rating scale technique to public school teaching is reported in the professional literature. (Price, 1979) Both service and business positions have been analyzed, but not in the public schools.

Let me begin my discussion of behaviorally anchored rating scales with a bit of their history. In 1963, Patricia Cain Smith and L. M. Kendall took note of the extreme demands placed upon the quality of ratings in many situations. They stressed that such ratings should possess reliability both across raters and situations. Both the levels and components of such ratings should be clearly understood by those assigned the task of making ratings. It is only under such circumstances that raters can make valid ratings, perhaps more originally, be expected to use rating scales with conviction or agreement. The necessary consensus among raters can be achieved only if the raters themselves define, in their own terms, the kind of behavior which represents each level of each discriminably different characteristic, and which dimension of behavior is illustrated by each kind of behavior. The behaviorally anchored rating scale is Smith and Kendall's answer to this problem.

In a few moments I shall show you several behaviorally anchored rating scales developed by Marianne Price in her doctoral study for the position of Special Education Teacher. I want first, however, to describe the procedure under which such scales are developed. Consideration of these two elements separately was inspired by Dickinson and Zellinger, 1980. It is through this procedure that the chief advantages of the behaviorally anchored rating scale are achieved. It may be noted in passing that the procedure for the development of behaviorally anchored rating scales has been studied informatively independent of the behaviorally anchored rating scale format itself.

Smith and Kendall describe the procedure for the development of behaviorally anchored rating scales as similar to that employed to ensure the faithfulness of translations from one language to another. Material is translated into a foreign language, and then retranslated by an independent translator into the original language. Where "slippage" is found to have occurred, corrections are made. You will see how this retranslation simile applies to the present context in a moment. The behaviorally anchored rating scale development procedure consists generally of the following five steps:

1. A representative sample of the persons who will ultimately serve as raters are enlisted to conduct the basic work of generating the behaviorally anchored rating scales.

In education, it would seem that this reference group could logically include not only administrative raters, but also teachers. The point of calling together the work group is not to generate administrative solidarity, but rather common understanding and commitment. Both sides of potential evaluation problems ought, in my judgment, to be involved in scale development from the outset. We have a long history in education of such involvement.

2. The group identifies and lists the broad qualities or characteristics to be evaluated. The most strongly supported dimensions are selected for further analysis. Usually the number of dimensions centers about eight. The participants' own terminology is retained in identifying the dimensions. Let me show you what such a set of dimensions might consist of.

SHOW OVERHEAD TRANSPARENCY #1. (RJB note the source)

3. The group then formulates general statements representing definitions of high, acceptable, and low performance for each dimension, and generates examples of behaviors at each level for each dimension. These are edited into the form of expectations of specific behavior. Let me show you an example of a set of behaviors indicative of one dimension.

SHOW OVERHEAD TRANSPARENCY #2. (RJB block off left columns.)

Note: It may be worth noting here that not all investigators have identified and defined the broad qualities, characteristics, or dimensions first. Some have begun with specific behaviors, which are then grouped into appropriate categories.

4. A second group of persons, representative of the group sampled by the original work group, is provided a listing of the broad dimensions, and a second list of the specific behaviors developed by the first group. They are asked to assign the behaviors to the dimensions. Behavioral examples are eliminated if there is not a criterion level of agreement on the dimension to which behaviors ought to be assigned. Qualities or dimensions are eliminated if there is not a criterion level of agreement in the behaviors which ought to be included.

5. Other judges, perhaps the first group, are asked to use the remaining behaviors to describe the behavior of satisfactory and unsatisfactory employees as across the several dimensions. A discrimination index is computed for each behavior. Judges are also asked to assign certain point values to each behavior within dimensions, according to its desirability. Items above a criterion level of point variability are eliminated. Mean point values for surviving behaviors are retained and used in the behaviorally anchored rating scale format.



Let me turn now to the scale format. You will recall the dimensions of performance I showed you a moment ago. Let me show you several of the scales representing these dimensions.

SHOW OVERHEAD TRANSPARENCIES #3, #4, and #5, if time permits.

The function of the scales is of course to assist in the evaluation of employee job performance. Let me describe how the rater is expected to perform this responsibility. Here I must confess to less guidance in the literature than has been present with respect to my prior comments. Nonetheless permit me to offer a few observations.

The rater is expected, over some no doubt specified period, to observe the employee's performance on the job. On each broad dimension of the appropriate set of rating scales, the rater assigns a point value to his or her rating of employee performance. The specific point values awarded are in accordance with the similarity between the behaviors observed and those for which points are specified on the behaviorally anchored rating scale. The specific comparison process followed, and this point is quite clear in Smith and Kendall's article, operates as follows. The rater forms a clear image of the employee's work, then predicts which of the behaviors listed in the scale the individual would be most likely to display.

7

It is here, in my judgment, that the behaviorally anchored rating is set adrift. The anchor drags in the mud of the rater's predictive powers. There is a second flaw, as I see it, which makes such anchors' value questionable. This is the permanent fixing of point values for the specific behaviors used as anchors. My experience both as ratee and rater suggests that few teaching behaviors maintain the same value from moment to moment, much less from teacher to teacher or class to class.

I recognize that the comments which I have just made sound contradictory. How can one complain both of the unreliability of point values in a scale, and then of the false precision in the scale? I suppose that the best defense is a good offense. I propose that we not sacrifice either for the other. A revision of the behaviorally anchored rating scale format, and corresponding minor changes in the scale development procedure would, I believe, preserve the best of both.

The specific behaviors generated to illustrate the dimensions agreed upon ought not to be sifted out on statistical criteria. The value of the behaviors is not in their statistical properties, but rather their representativeness of actual teaching behavior. The larger number of behaviors generated could serve as a guide to observation, and could provide behavioral anchors, but not for numerical ratings, but rather for the exercise of judgment on the part

of the rater. Perhaps some dichotomization of positive and negative behaviors could be performed to simplify the job of rating, but this step is beside the point which I am trying to make.

The behaviorally anchored rating scale is not a free lunch. Yet it may provide enough nourishment for us to continue the search productively for improved methods of teacher evaluation.

## REFERENCES

- Cummings, L. L. and Donald P. Schwab. Performance in Organizations. Glenview, IL: Scott, Foresman & Co., 1973.
- De Cotiis, Thomas A. "A Critique and Suggested Revision of Behaviorally Anchored Rating Scales Development Procedures" Educational and Psychological Measurement 38 (3), Autumn 1978, 681-90.
- Dickinson, Terry L. and Peter M. Zellinger. "A Comparison of the Behaviorally Anchored Rating and Mixed Standard Scale Formats." Journal of Applied Psychology 65 (2), April, 1980, 147-54.
- Price, Marianne. "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Specific Education Teachers." Ed.D. Dissertation, Temple University, 1979.
- Schneier, Craig Eric and Richard W. Beatty. "Developing Behaviorally-Anchored Rating Scales (BARS)" The Personnel Administrator 24 (8), August, 1979, 59-68.
- Schwab, Donald P. and Herbert G. Heneman III. "Behaviorally Anchored Rating Scales: A Review of the Literature." Personnel Psychology 28 (4), Winter, 1975, 549-62.
- Smith, Patricia Cain and L. M. Kendall. "Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales." Journal of Applied Psychology 47 (2), April 1963, 149-55.

BEHAVIORALLY ANCHORED RATING SCALES FOR EVALUATING  
THE SPECIAL EDUCATION TEACHERS OF THE  
MONTGOMERY COUNTY, PENNSYLVANIA,  
INTERMEDIATE UNIT

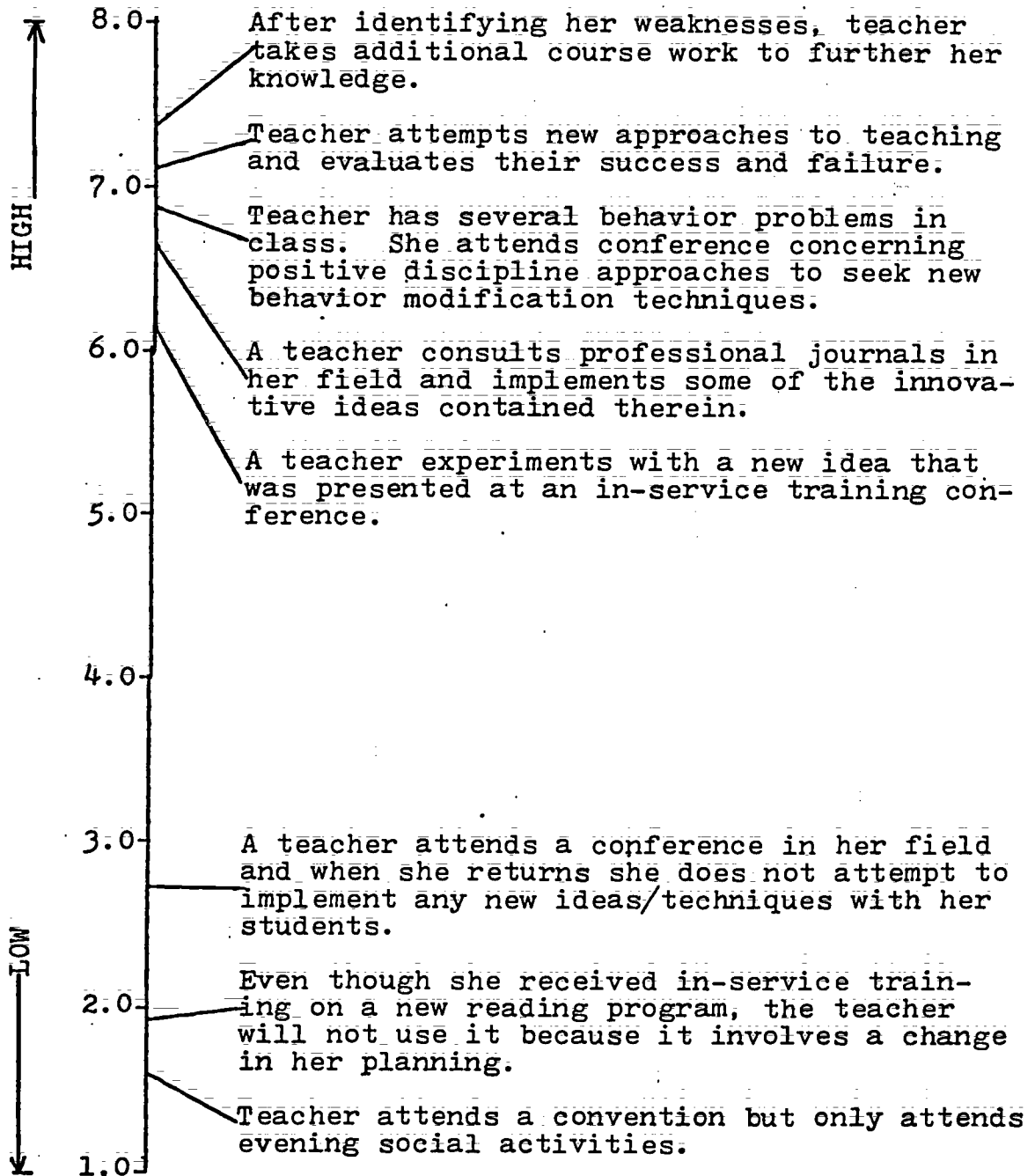
1. PROFESSIONAL GROWTH
2. COMMUNICATION SKILLS
3. RAPPORT WITH STUDENTS
4. APPROPRIATE PLANNING
5. INSTRUCTIONAL SKILLS
6. RECORD KEEPING AND REPORTING
7. BEHAVIOR MANAGEMENT
8. INTERSTAFF RELATIONSHIPS

DEVELOPED BY MARIANNE PRICE

1978

## 1. PROFESSIONAL GROWTH

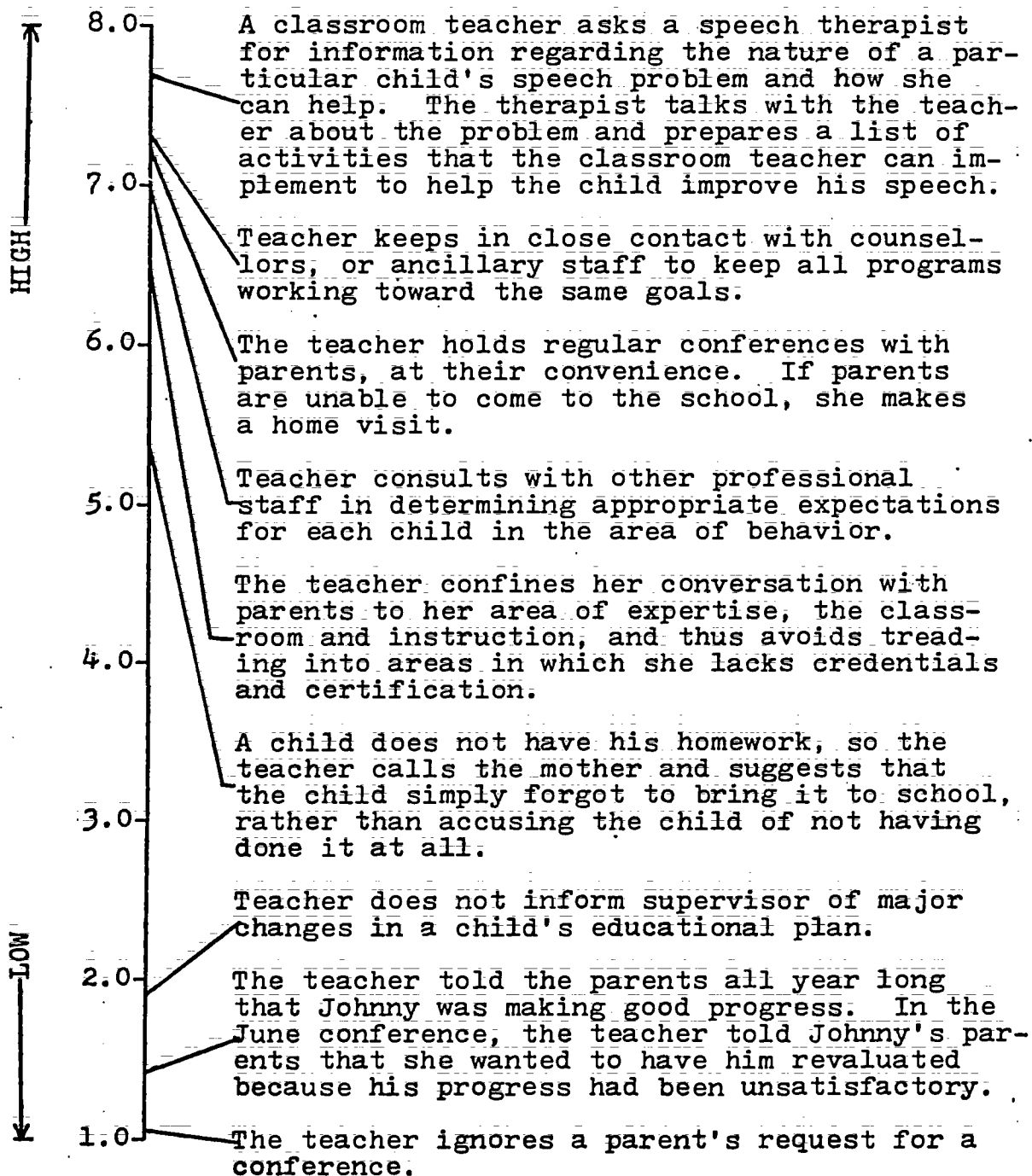
Seeks to improve instruction  
through the exploration of suggestions,  
new ideas, materials, approaches, etc.



Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University, 1978), p. 246. Reproduced by permission.

## 2. COMMUNICATION SKILLS

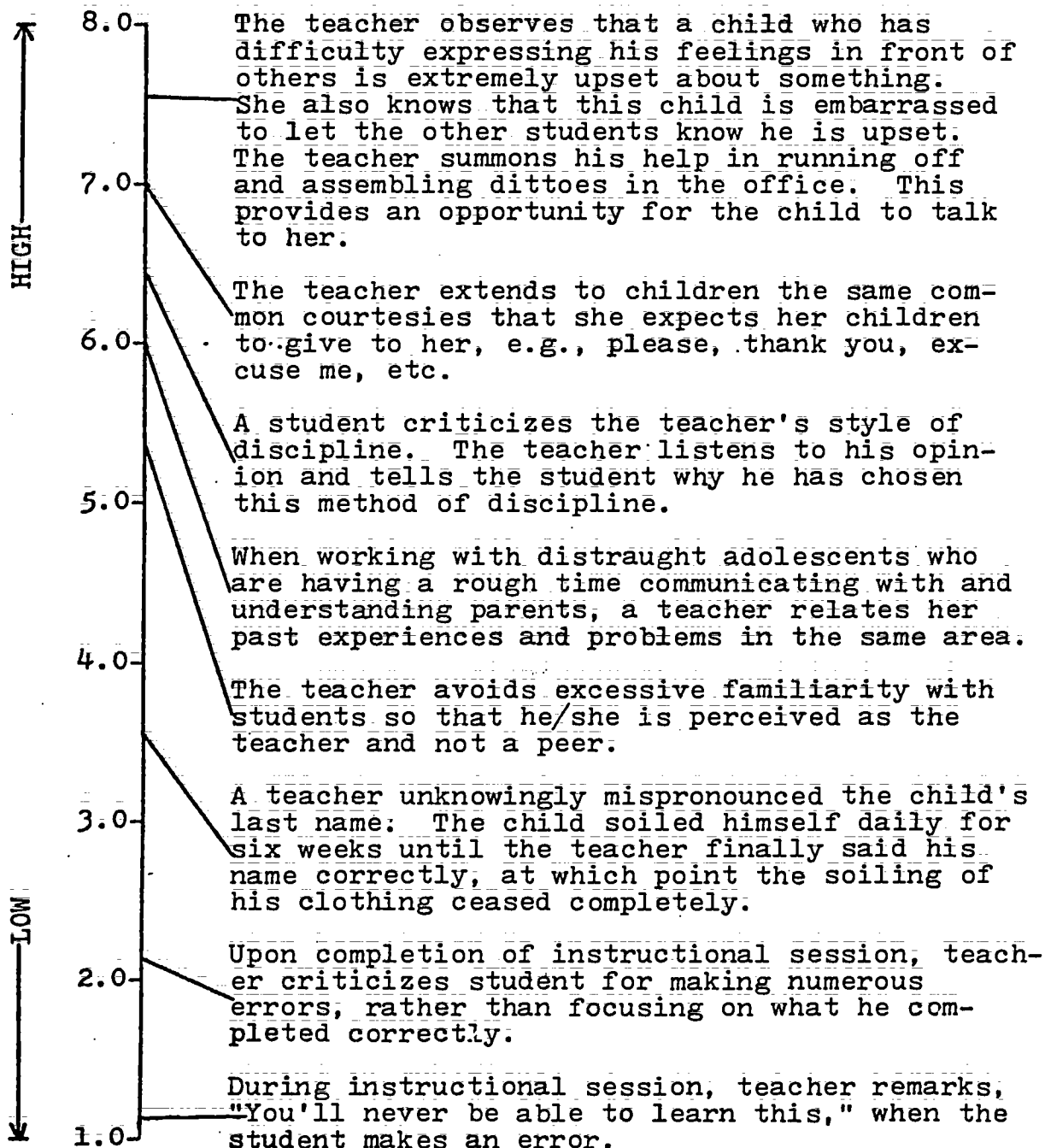
The ability to exchange information with administrators, professional persons, parents, and other adults



Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University, 1978), p. 247. Reproduced by permission.

### 3. RAPPORT WITH STUDENTS

The ability to develop an accepting relationship with children, which facilitates their success and growth in a learning environment.



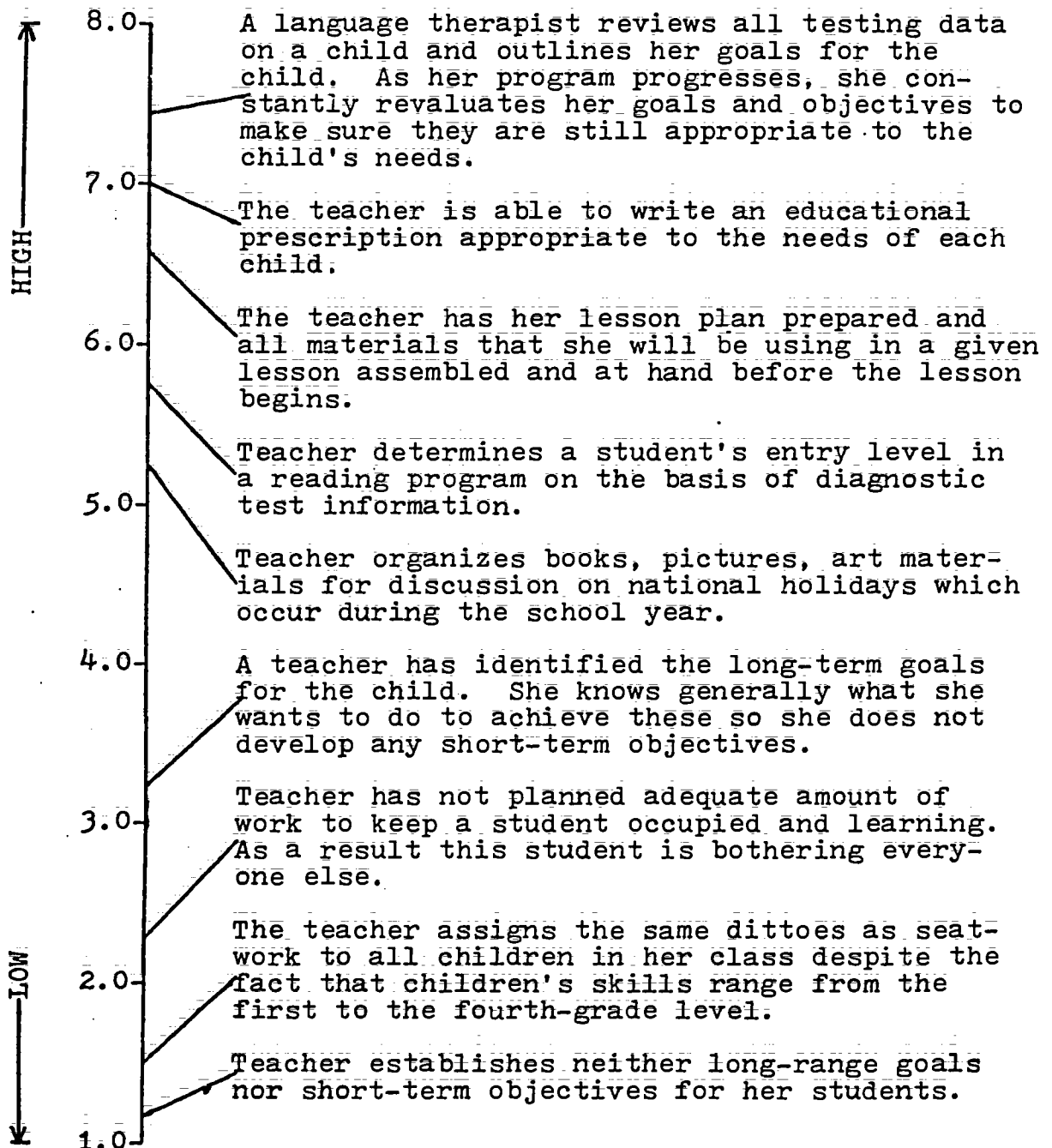
Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University, 1978), p. 248. Reproduced by permission.



#### 4. APPROPRIATE PLANNING

The ability to organize instruction in order to achieve short and long-term objectives based on the needs and level of performance of each child.

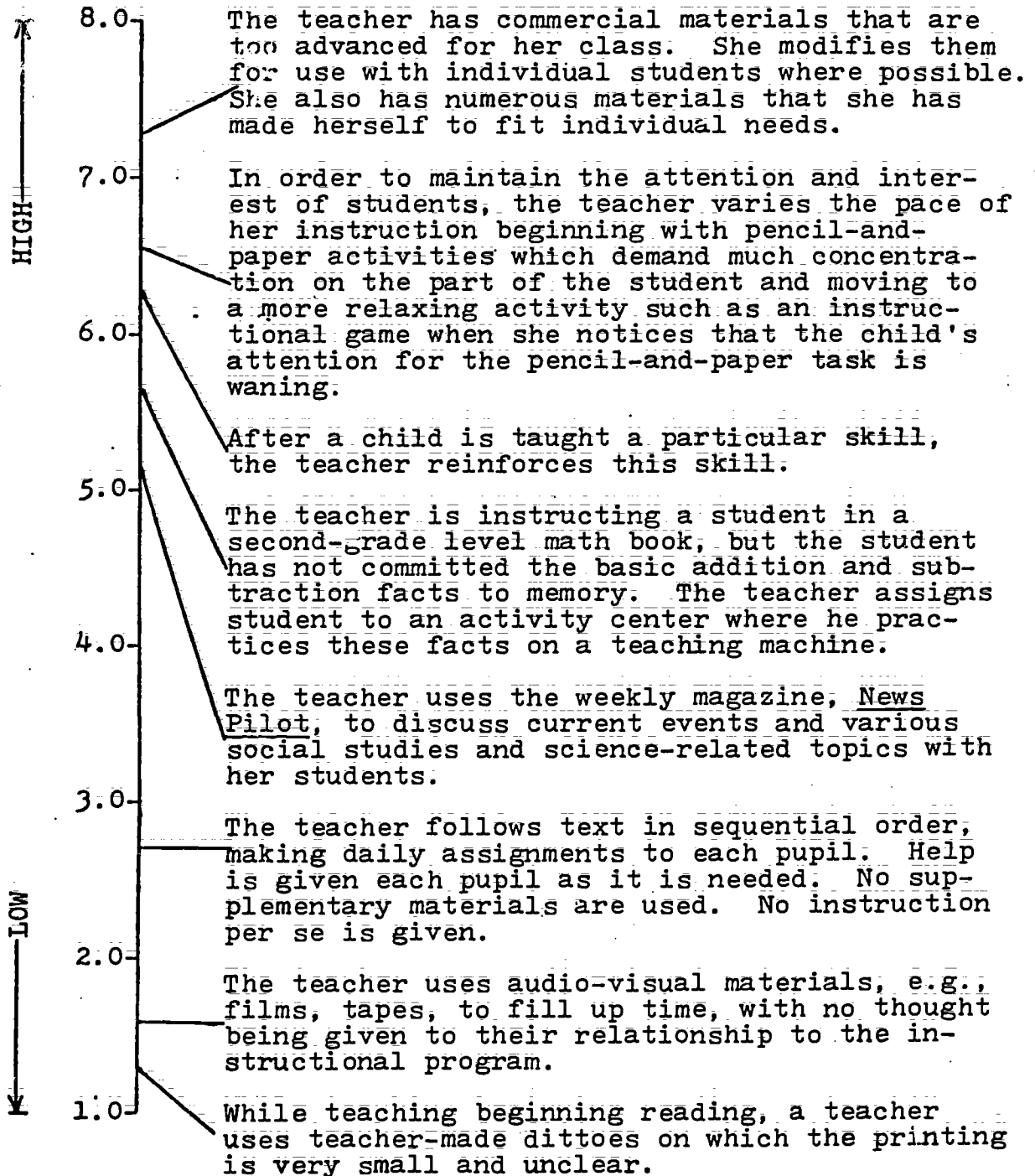
---



Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University, 1978), p. 249. Reproduced by permission.

## 5. INSTRUCTIONAL SKILLS

The ability to implement the educational program through the use of appropriate materials and techniques.

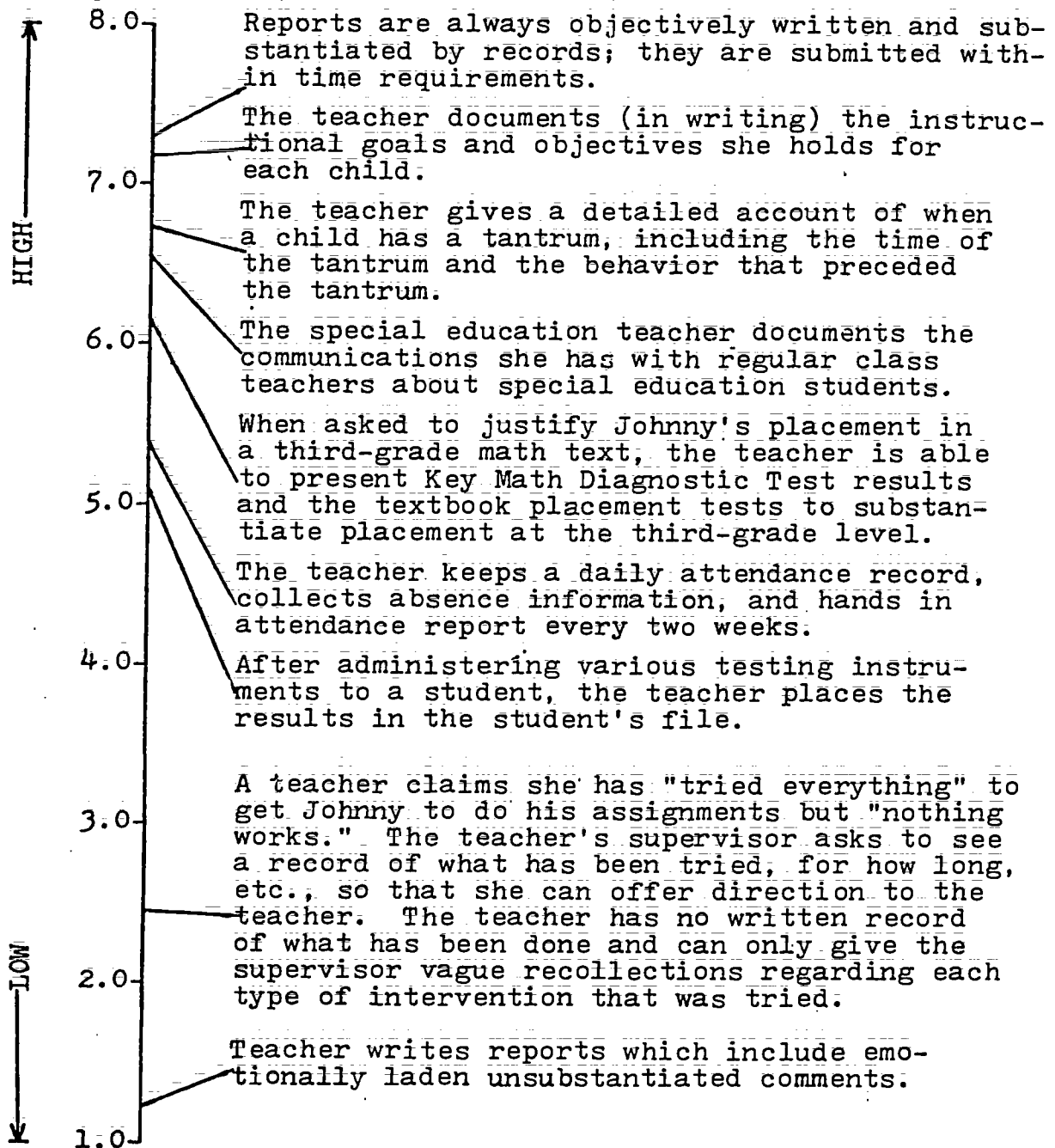


Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University. 1978). p. 250. Reproduced by permission.

## 6. RECORD KEEPING AND REPORTING

The ability to keep accurate and current records, using this information for comprehensive reporting as required.

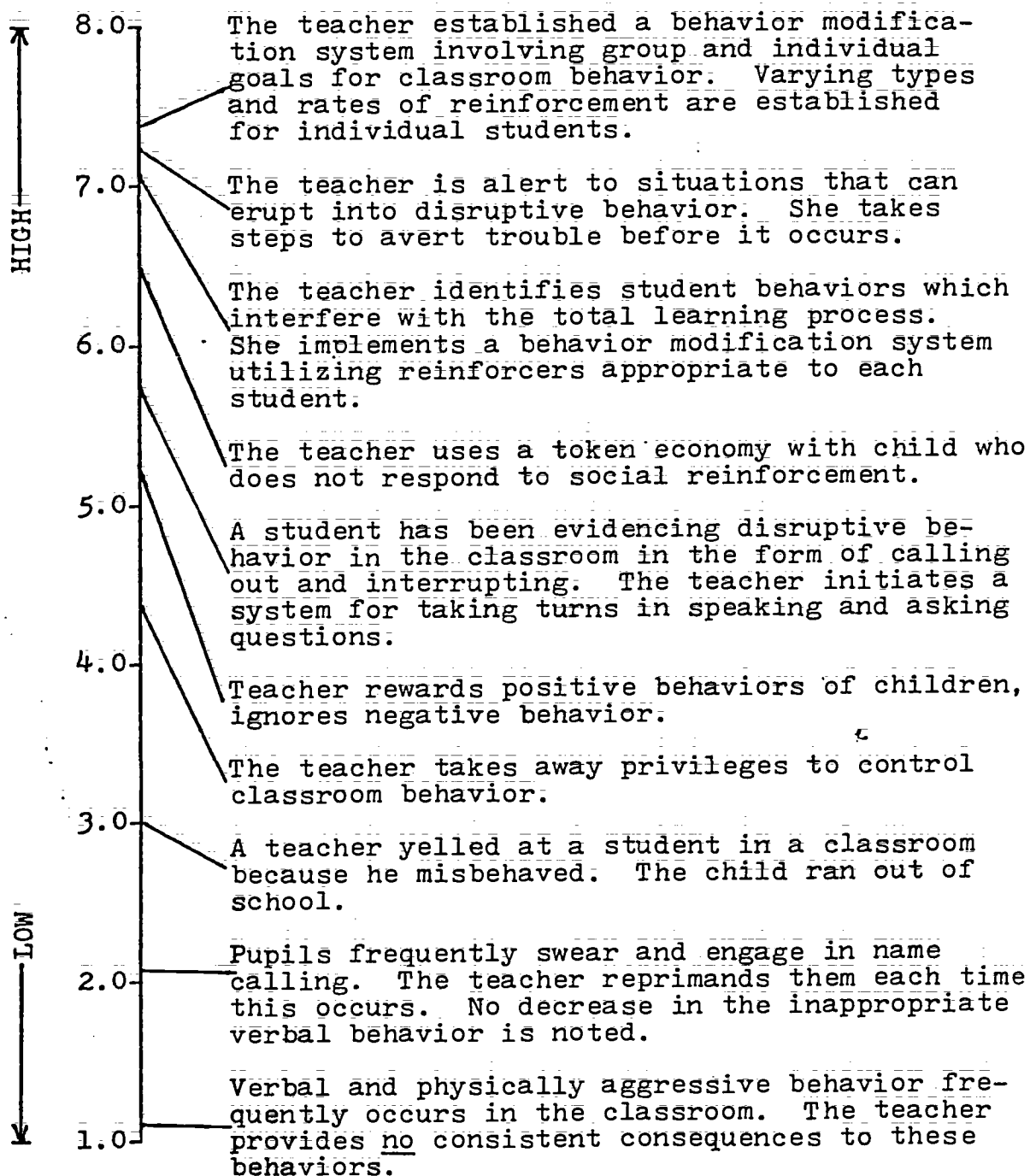
---



Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University, 1978), p. 251. Reproduced by permission.

## 7. BEHAVIOR MANAGEMENT

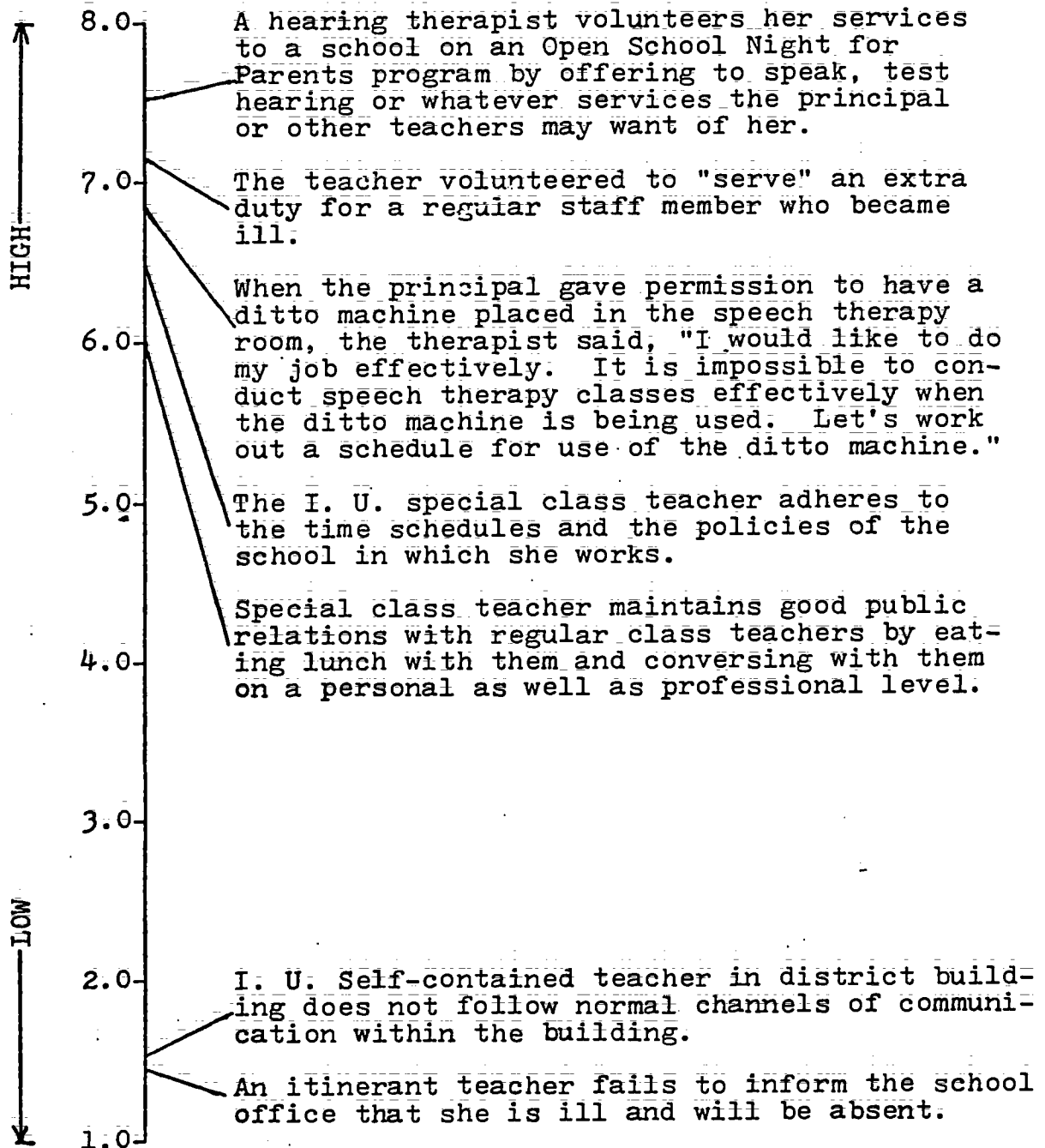
The ability to establish and maintain a behavioral atmosphere that is conducive to learning.



Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University, 1978), p. 252. Reproduced by permission.

## 8. INTERSTAFF RELATIONSHIPS

The ability to build and develop positive relationships with I. U. and/or district staff.



Source: Marianne Price, "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Special Education Teachers" (EdD dissertation, Temple University, 1978), p. 253. Reproduced by permission.